

(19) 世界知的所有権機関
国際事務局



(43) 国際公開日
2005年1月13日 (13.01.2005)

PCT

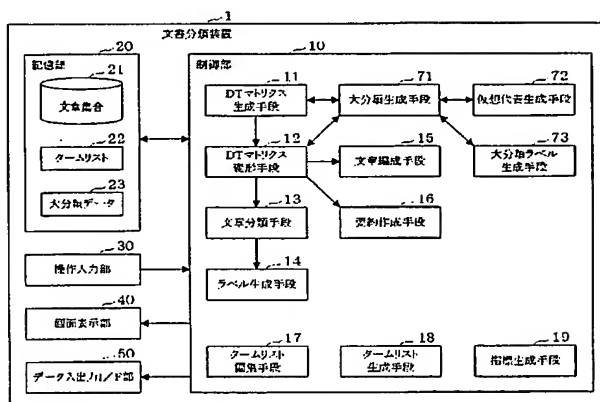
(10) 国際公開番号
WO 2005/004004 A1

- (51) 国際特許分類⁷: G06F 17/30 (72) 発明者; および
(21) 国際出願番号: PCT/JP2004/009341 (75) 発明者/出願人(米国についてのみ): 村上 英治 (MURAKAMI, Eiji) [JP/JP]; 〒1508316 東京都渋谷区渋谷2丁目12番19号 株式会社 山武内 Tokyo (JP). 寺野 隆雄 (TERANO, Takao) [JP/JP]; 〒1680063 東京都杉並区和泉3丁目15番16号 Tokyo (JP).
(22) 国際出願日: 2004年7月1日 (01.07.2004)
(25) 国際出願の言語: 日本語
(26) 国際公開の言語: 日本語
(30) 優先権データ:
特願2003-189716 2003年7月1日 (01.07.2003) JP
特願2004-053419 2004年2月27日 (27.02.2004) JP
(71) 出願人(米国を除く全ての指定国について): 株式会社 山武 (YAMATAKE CORPORATION) [JP/JP]; 〒1508316 東京都渋谷区渋谷2丁目12番19号 Tokyo (JP).
(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG,

[続葉有]

(54) Title: SENTENCE CLASSIFICATION DEVICE AND METHOD

(54) 発明の名称: 文章分類装置および方法



- 1... SENTENCE CLASSIFICATION DEVICE
20... STORAGE SECTION
21... SENTENCE SET
22... TERM LIST
23... HIGHER-LEVEL CLASSIFICATION DATA
30... OPERATION INPUT SECTION
40... SCREEN DISPLAY SECTION
50... DATA I/O SECTION
10... CONTROL SECTION
11... DT MATRIX GENERATION MEANS
12... DT MATRIX DEFORMATION MEANS
13... SENTENCE CLASSIFICATION MEANS
14... LABEL GENERATION MEANS
71... HIGHER-LEVEL CLASSIFICATION GENERATION MEANS
72... VIRTUAL REPRESENTATIVE GENERATION MEANS
15... SENTENCE EDITING MEANS
73... HIGHER-LEVEL CLASSIFICATION LABEL GENERATION MEANS
16... SUMMARY GENERATION MEANS
17... TERM LIST EDITING MEANS
18... TERM LIST GENERATION MEANS
19... INDEX GENERATION MEANS

(57) Abstract: DT matrix generation means (11) generates a DT matrix (11A) from a sentence D in a sentence set (21) and a term T in a term list (22). DT matrix deformation means (12) DM-decomposes the DT matrix (11A) and generates a deformed DT matrix (11B). For each cluster appearing on the deformed DT matrix (11B), sentence classification means (13) extracts each sentence (D) belonging to the cluster as one class (subset).

(57) 要約: DTマトリクス生成手段11で、文章集合21内の各文章Dとタームリスト22内の各タームTとからDTマトリクス11Aを生成し、DTマトリクス変形手段12で、そのDTマトリクス11AをDM分解して変形DTマトリクス11Bを生成する。そして文章分類手段13で、変形DTマトリクス11B上に現れる各クラスごとに、当該クラスに属する各文章Dを1つの分類(部分集合)として抽出出力する。

BEST AVAILABLE COPY

WO 2005/004004 A1



SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ,
VC, VN, YU, ZA, ZM, ZW.

BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN,
TD, TG).

(84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF,

添付公開書類:

— 国際調査報告書

2文字コード及び他の略語については、定期発行される各PCTガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

BEST AVAILABLE COPY